КОЛИЧЕСТВЕННЫЕ ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ ИСТОЧНИКОВ СООБЩЕНИЙ И КАНАЛОВ

Классификации подходов к оценке количества информации

Количество информации в дискретном сообщении.

Синтаксические меры информации

Избыточность источника дискретных сообщений

Количественные информационные оценки дискретных источников с памятью

- условная и взаимная информация
- совместная и условная энтропия

Производительность источника дискретных сообщений

Пропускная способность дискретного канала

Классификации подходов к оценке количества информации

Синтаксическая мера количества информации оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту (учитываются скорость передачи, размеры кодов представления информации).

Семантическая мера информации используется для измерения смыслового содержания информации. Связана с понятием тезауруса.

Прагматическая мера определяет полезность информации (ценность) для достижения пользователем поставленной цепи.

Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Виды источников информации:

- дискретный;
- комбинаторный;
- вероятностный;
- марковский;
- бернуллиевский.

Дискретный ансамбль:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p(x_1) & p(x_2) & \dots & p(x_N) \end{pmatrix}, \ \sum_{i=1}^{N} p(x_i) = 1$$

Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Требования к вводимой мере оценки количества информации:

1) Чем больше число возможных сообщений (возможных значений сигнала), тем больше априорная неопределенность и тем большее количество информации получает адресат, когда эта неопределенность снимается. Если же выбор сообщения заранее предопределен, то количество информации в этом сообщении равно нулю.

2) Вводимая мера должна обладать свойством аддитивности, в соответствии с которым неопределенность объединенного источника равна сумме неопределенностей исходных источников.

Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Комбинаторный подход к оценке количества информации (Р. Хартли, 1928г.).

Степень неопределенности опыта X с N различными исходами характеризуется числом

$$H(X) = \log N.$$

Не учитываются вероятности различных исходов.

Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Вероятностный подход к оценке количества информации (К. Шеннон, 1949г.).

Степень неопределенности конкретного состояния зависит не только от объема алфавита источника, но и от вероятности этого состояния.

Количество информации, содержащееся в одном элементарном дискретном сообщении x_k целесообразно определить как функцию вероятности появления этого сообщения $p(x_k)$ и характеризовать величиной

$$i(x_k) = -\log p\left(x_k\right) = \log \frac{1}{p(x_k)}$$

Величина $i(x_k)$ называется количеством собственной информации в сообщении $x_k \in X$.

Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Вероятностный подход к оценке количества информации (К. Шеннон, 1949г.).

Для цифровой характеристики всего ансамбля или источника сообщений используется математическое ожидание количества информации в отдельных сообщениях, называемое энтропией:

$$H(X) = M\{i(x)\} = \sum_{X} i(x)\,p(x) = -\sum_{X} p(x)\log p(x)$$

Энтропия представляет собой среднее количество собственной информации в сообщениях дискретного источника без памяти.

Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Свойства энтропии

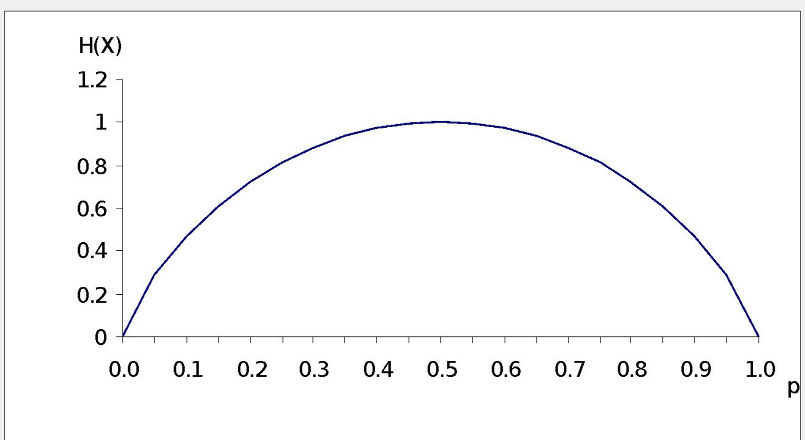
1) Энтропия всякого дискретного ансамбля неотрицательна $H(X) \geq 0$.

2) Пусть N – число сообщений в ансамбле. Тогда $H(X) \leq \log N$.

3) Энтропия обладает свойством аддитивности.

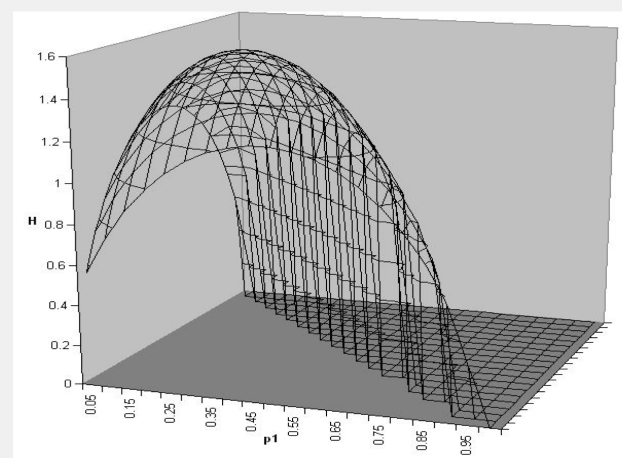
Энтропия двоичного источника без памяти:

$$H(X) = -\left[p \log_2(p) + (1-p)\log_2(1-p)\right]$$



Энтропия троичного источника без памяти:

$$H(X) = -\left[\begin{array}{l} p_1 \log_2 p_1 + p_2 \log p_2 + \\ + (1 - p_1 - p_2)\log_2(1 - p_1 - p_2) \end{array}\right]$$



Количество информации в дискретном сообщении. Синтаксические меры оценки количества информации

Алгоритмический подход к оценке количества информации (А.Н. Колмогоров, 1965г.).

Энтропия $H(X, Y)$ ("колмогоровская сложность" объекта Y при заданном X) есть мнимая длина, записанная в виде последовательности нулей и единиц, программы, которая позволяет построить объект *Y*, имея в своем распоряжении объект *X.*

Колмогоровская сложность обычно невычислима.

Избыточность источника дискретных сообщений

Максимальную энтропию имеет источник, все сообщения которого передаются равновероятно и независимо.

Невыполнение этих требований приводит к уменьшению энтропии и появлению избыточности источника.

Понятие избыточности источника сообщений связано с мощностью алфавита источника и его энтропией:

$$\chi = \frac{\log N - H(X)}{\log N} = 1 - \frac{H(X)}{\log N}$$

При $\chi=0$ источник называют источником без избыточности

Количественные информационные оценки дискретных источников с памятью

Источник сообщения обладает памятью, если между элементами сообщения одного или нескольких источников имеется детерминированная или статистическая связь.

Сообщения, вырабатываемые таким источником – сложные сообщения.

При определении количества информации в таких сообщения необходимо учитывать условные вероятности появления элементарных сообщений.

Количественные информационные оценки дискретных источников с памятью. Условная и взаимная информация

Пусть $\{XY, p(x_i,y_j)\}$ – два совместно заданных ансамбля $\{X, p(x_i)\}$ и $\{Y, p(y_j)\}$.

Зафиксируем некоторое сообщение y_j и рассмотрим условное распределение на *X.*

Апостериорная вероятность $p(x_i|y_j)$ - неопределенность, остающаяся о сообщении x_i после того, как было принято сообщение y_j.

Условная собственная информация:

$$i(x_i \mid y_j) = -\log p(x_i \mid y_j)$$

Совместная информация пары событий:

$$i(x_i, y_j) = i(x_i) + i(y_j) - \log\frac{p(x_i \mid y_j)}{p(x_i)} = i(y_j) + i(x_i) - \log\frac{p(y_j \mid x_i)}{p(y_j)}$$

Взаимная информация пары событий:

$$i(x_i; y_j) = \log\frac{p(x_i \mid y_j)}{p(x_i)} = \log\frac{p(y_j \mid x_i)}{p(y_j)}$$

Количественные информационные оценки дискретных источников с памятью. Совместная и условная энтропия

Для характеристики всего ансамбля принято использовать математические ожидания случайных величин.

Энтропия (совместная энтропия) ансамбля *XY*:

$$H(X,Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j)\log p(x_i, y_j)$$

$$H(X,Y) = \underline{-\sum_{j=1}^{M} p(y_j)\log p(y_j)} - \sum_{j=1}^{M} p(y_j)\underline{\sum_{i=1}^{N} p(x_i \mid y_j)\log p(x_i \mid y_j)}$$

$$H(X,Y) = H(Y) + \left(-\sum_{j=1}^{M} p(y_j) H(X \mid y_j)\right)$$

Сумма в скобках - условная энтропия источника *X* относительно источника *Y*, обозначается как *H*(*X* | *Y*).

$H(X, Y) = H(Y) + H(X \mid Y)$ или $H(X, Y) = H(X) + H(Y \mid X)$

Количественные информационные оценки дискретных источников с памятью. Средняя взаимная информация

Математическое ожидание случайной величины $i(x_i;y_j)$ - *средняя взаимная информация* между источниками X и Y:

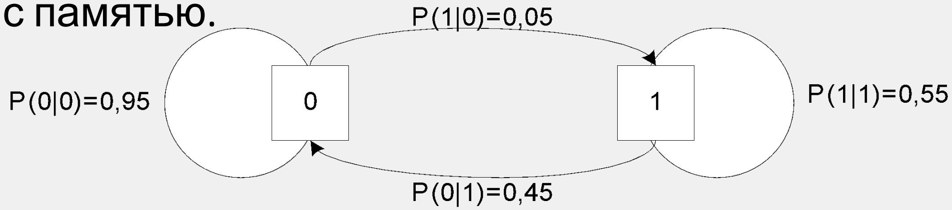
$$I(X;Y)=\sum_{i=1}^{N}\sum_{j=1}^{M}p(x_i,y_j)i(x_i;y_j)=\sum_{i=1}^{N}\sum_{j=1}^{N}p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)p(y_j)}$$

можно записать:

$$I(X;Y)=H(X)-H(X\mid Y)=H(Y)-H(Y\mid X)$$

Количественные информационные оценки дискретных источников с памятью. Оценка информативности источников с памятью

Пример. Энтропия двоичного источника с памятью. Дан двоичный (двухсимвольный) Марковский источник, определенный вероятностями переходов состояний p(0|1)=0,45 и p(1|0)=0,05. Найти энтропию источника с памятью.



Энтропия источника:

$$H(X) = p(0)H(X\,|\,0) + p(1)H(X\,|\,1)$$

где:

$$H(X\,|\,0) = -\big[p(0\,|\,0)\log_2 p(0\,|\,0) + p(1\,|\,0)\log_2 p(1\,|\,0)\big] = 0{,}286$$

$$H(X\,|\,1) = -\big[p(0\,|\,1)\log_2 p(0\,|\,1) + p(1\,|\,1)\log_2 p(1\,|\,1)\big] = 0{,}993$$

Априорная вероятность каждого состояния находится либо итерационным перемножением матрицы переходов, либо с помощью системы линейных уравнений:

$$p(0) = p(0\,|\,0)\,p(0) + p(0\,|\,1)\,p(1)$$

$$p(1) = p(1\,|\,0)\,p(0) + p(1\,|\,1)\,p(1)$$

$$p(0) + p(1) = 1$$

Решая ее, находим $p(0)$=0,9 и $p(1)$=0,1.

Энтропия источника без памяти H(X) = -(p(0)*log p(0) + p(1) * log p(1)) = 0,469 бит/символ.

Энтропия источника с памятью:

$$H(X) = p(0)H(X\,|\,0) + p(1)H(X\,|\,1) = 0{,}9 \cdot 0{,}286 + 0{,}1 \cdot 0{,}993 = 0{,}357\,\textit{бит}\,/\,\textit{символ}$$

Производительность источника дискретных сообщений

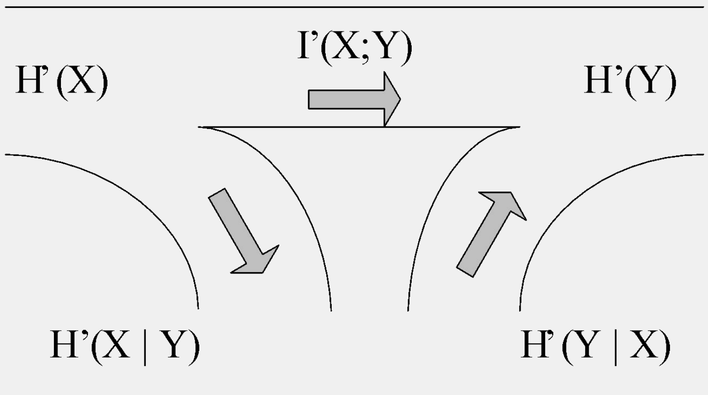
Производительность источника $H'(X)$ - суммарная энтропия сообщений, переданных за единицу времени (бит/сек)

$$H'(X) = \frac{1}{T} H(X)$$

Аналогично для условной энтропии и количества информации в единицу времени

$$H'(X \mid Y) = \frac{1}{T} H(X \mid Y) \qquad I'(X,Y) = \frac{1}{T} I(X,Y)$$

Если X – ансамбль сигналов на входу дискретного канала, а Y – ансамбль сигналов на его выходе, то скорость передачи информации по каналу $I'(X,Y) = H'(X) - H'(X \mid Y) = H'(Y) - H'(Y \mid X)$



Пропускная способность дискретного канала

Максимальное количество переданной информации, взятое по всевозможным источникам входного сигнала, характеризует сам канал и называется пропускной способностью канала связи в расчете на один символ (бит/символ):

$$C_{символ} = \max_{p(X)} I(X,Y)$$

Пропускная способность канала в расчете на единицу времени (бит/сек):

$$C = \max_{p(X)} I'(X,Y)$$